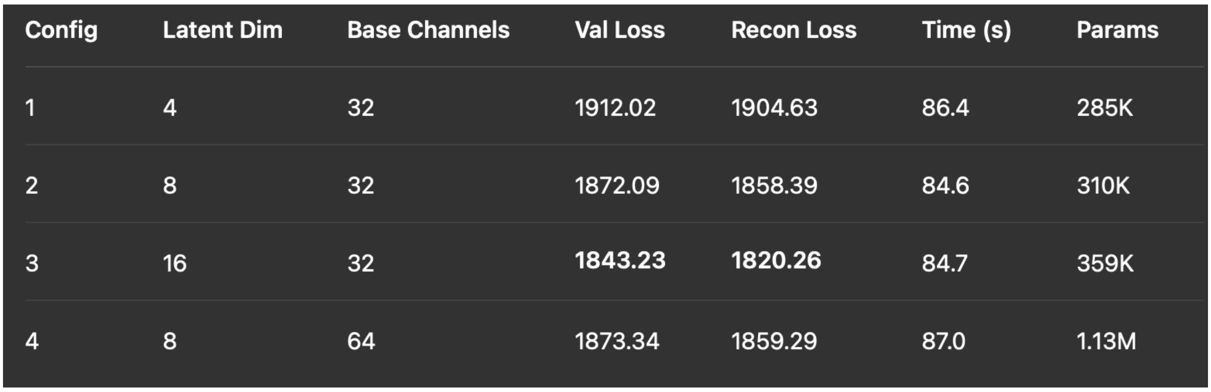
Assignment 7 Report: Autoencoders and VAEs on CIFAR-10

B.1: Architecture Configuration Comparison

We trained four VAE configurations on CIFAR-10 to study how latent size and network capacity affect reconstruction quality and efficiency. *All models were trained for 10 epochs (β = 1.0, LR = 3e-3).*



Config	Latent Dim	Base Channels	Val Loss	Recon Loss	Time (s)	Params
1	4	32	1912.02	1904.63	86.4	285K
2	8	32	1872.09	1858.39	84.6	310K
3	16	32	**1843.23**	**1820.26**	84.7	359K
4	8	64	1873.34	1859.29	87.0	1.13M

Performance Analysis

- **Config 1 (4D)**: Small latent space led to the poorest reconstructions and highest loss due to severe information bottleneck.
- **Config 2 (8D)**: Balanced performance and efficiency; serves as baseline.
- **Config 3 (16D)**: Best overall, achieving the lowest losses with minimal added cost. The latent dimension effectively captures complex visual details without overfitting.
- **Config 4 (Deeper)**: Over-parameterized; higher loss despite 3× more parameters, suggesting redundancy and mild overfitting.

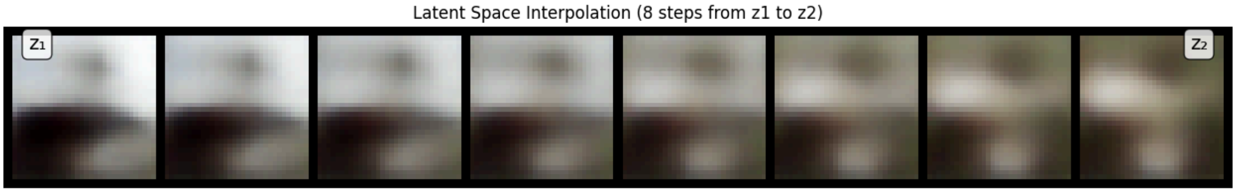
B.2: Final Architecture Selection

Selected Configuration: Config 3 (Large Latent, 16D)

Justification

- **Highest quality:** Lowest reconstruction loss (1820.26) — 4.4% better than Config 2.
- **Efficient:** Only 16% more parameters yet identical training time.
- **Stable & well-regularized:** Smooth KL divergence (≈25) and no mode collapse.
- **Optimal compression:** 3072× reduction (3×32×32 → 16) while preserving visual fidelity.

B.3 Latent Space Interpolation



Latent Space Interpolation (8 steps from z1 to z2)

Figure 1: Linear interpolation between two latent vectors (Config 3).

Interpolating between two random latent vectors produces **smooth, semantically coherent transitions** between images. Intermediate reconstructions blend colors, shapes, and textures gradually, confirming that the learned latent space is **continuous and structured**. The absence of artifacts shows that the decoder consistently maps nearby points to visually similar outputs—demonstrating the VAE's ability to learn a meaningful, disentangled latent representation.